Primer on Measure Theory and Probability

David Matthews

November 12, 2013

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of a sample space Ω, a σ-algebra $\mathcal{F}$ (the set of all possible measureable events $E \in \mathcal{F}$) and a probability measure $\mathbb{P}$ where $\mathbb{P}(E)$ is the probability that event E occurs. In particular we will see that a random variable in probability theory corresponds to a measurable function in measure theory. Furthermore the integral of that function over the whole probability space (if the random variable is absolutely convergent) is known as the expectation of that random variable.

1 The de Finetti notation

Let $\mathcal{X}$ be some set and $A \subseteq \mathcal{X}$, then we can define an indicator function $\mathbb{I}_A$ such that $\mathbb{I}_A(X) = 1$ if $X \in \mathcal{A}$ and 0 otherwise. There is a correspondence between A and $\mathbb{I}_A$ so we write $\tilde{A} := \mathbb{I}_A$.

Now recall that a probability measure $\mathbb{P}$ really denotes a map from some sigma field $\mathcal{A}$ of subsets of state space Ω to the interval $[0, 1]$. There exists a correspondence between the sigma algebra $\mathcal{A}$ and a subset of the collection of random variables on Ω. Since the expectation map maps random variables to $\mathbb{R}$ we can write $\mathbb{E}(\tilde{A}) = \mathbb{P}(A))$. Therefore instead of writing $\mathbb{E}X$ for expectation of some event X we use the notation $\mathbb{P}X$. This has the advantage that given two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on some state space Ω we can use $\mathbb{P}$ for the expectation corresponding to $\mathbb{P}$ etc.

2 Random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and X be a random variable then X is a real valued function from the statespace Ω to $\mathbb{R}$. In particular X takes the pair $(X, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ where $\mathcal{B}(\mathbb{R})$ is the Borel field on the reals.

In general if $\mathcal{X}$ is any set with sigma-field $\mathcal{A}$ and $\mathcal{Y}$ is a set with a sigma-field $\mathcal{B}$. Now T be a function $T : \mathcal{X} \to \mathcal{Y}$. We say that T is $\mathcal{A} \setminus \mathcal{B}$-measurable if the preimage $\{X \in \mathcal{X} : TX \in B\}$ belongs to $\mathcal{A}$ for every $B \in \mathcal{B}$. Now suppose that μ is a measure on the space $\mathcal{X}$ and that there exists a $\mathcal{A} \setminus \mathcal{B}$-measurable function T. We can construct the *image measure*, ν, on $\mathcal{B}$ by setting

$$\nu B := \mu(T^{-1}B),$$

for any $b \in \mathcal{B}$.

Exercise. Prove that ν derived in this way is a measure.

Exercise. Now consider the case when X is a random variable, i.e. a map from $(\Omega, \mathcal{F}, \mathbb{P})$ to $\mathcal{B}(\mathbb{R}$. Check that X is an $\Omega \backslash \mathbb{R}$ measurable function.

The image measure of $\mathbb{P}$ in the case when X is a random variable is called $X(\mathbb{P})$ and it is often written as $\mathbb{P}_X$ and it is called the *distribution* of X. The distribution of a variable will be the key to our later definitions.

2.1 Strong law for large numbers

The strong law for large numbers (SLLN) asserts that a collection of averages converges almost surely to the expectation. In its most common form the (SLLN) is stated a s below.

Theorem 1 (SLLN (Kolmogorov))**.** *Let $X_1, X_2, \dots$ be independent, integrable random variables with the same distributions and an expectation, μ,, then the average, $\frac{(X_1+\cdots+X_n)}{N}$ converges almost surely to μ.*

For example if we take sequences of coin flips the proportion of heads after n flips will almost surely converge to $\frac{1}{2}$ as n approaches infinity.

2.2 Product Measures

We might ask: does a countably infinite sequence of random variables, $\{X_i : i \in \mathbb{N}\}$ necessarily exist? In the finite case, if we have some finite n random variables $X_1, X_2 \dots, X_n$ such that each X_i is a map from a probability space $(\Omega_\mathrm{n}, \mathcal{F}_\mathrm{n}, \mathbb{P}_\mathrm{n})$ we can always construct a probability measure $\mathbb{P}^n$ on the product space $\Omega^n = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n$ with the product Borel field $\mathcal{F}^n = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n$.

We can ask for two product measure $\mathbb{P}^n$ and $\mathbb{P}^{n+1}$ to be consistent in the expected way then we get a lovely corollary.

Corollary 1. *For probability measures P_i on arbitrary measure spaces $(\Omega_\mathrm{n}, \mathcal{F}_\mathrm{n}, \mathbb{P}_\mathrm{n})$ there exists a probability measure $\mathbb{P}$ such that*

$$\mathbb{P}(\mathcal{F}_1 \times \mathcal{F}_2 \times \cdots \times \mathcal{F}_n) = \Pi_{i \le k} P_i A_i$$

for all measurable rectangles.

2.3 Convergence in Distribution

A real-valued function f on a metric space $\mathcal{X}$ is said to be Lipschitz if there exists finite K such that for all $x, y \in \mathcal{X}$:

$$|f(x) - f(y)| \le K d(x, y).$$

We call the space of bounded Lipschitz functions on $\mathcal{X}$, $BL(\mathcal{X})$.

Definition 2.1. A sequence of probability measure $\{P_n\}$ on $\mathcal{B}(\mathcal{X})$ converge weakly to a probability measure P on $\mathcal{B}(\mathcal{X})$ if $P_n f \to Pf$ for all $f \in BL(\mathcal{X})$.

Let $\mathcal{X}$ be a metric space with metric $d(\ ,\ )$ equipped with its Borel σ-field $\mathcal{B}(\mathcal{X})$. Recall that a random element $X \in \mathcal{X}$ is a $\mathcal{F} \backslash \mathcal{B}(\mathcal{X})$-measurable map from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ into $\mathcal{X}$ and that the image measure $X(\mathbb{P}$ is called the distribution of $\mathcal{X}$ under $\mathbb{P}$. Now we can define the convergence in distribution of random elements $X_1, X_2, \dots$ to a probability measure P

We can define the convergence in distribution of a sequence of random events $X_1, X_2, \dots,$ to a probability measure P on $\mathcal{B}(\mathcal{X})$ to be if their distributions converge weakly to P.

Remark 1. *Convergence of distribution is not related to point wise convergence of the X_n as functions. Indeed, every X_n might be defined on a different probability space $(\Omega_\mathrm{n}, \mathcal{F}_\mathrm{n}, \mathbb{P}_\mathrm{n})$.*

3 Janson's Paper

Let X_{ni} be the number of vertices of outdegree $i \geq o$ in a random recursive tree with n vertices.

Theorem 3.1 (Janson 1.1)**.** *(i) As* $n \to \infty, n^{-1}X_{ni} \to 2^{-i-1}$ *almost surely, and*

(ii) $n^{\frac{1}{2}}(X_{ni} - 2^{-i-1}n) \to (d)V_i$

Where V_i *are jointly Gaussian variables with means* $\mathbb{E}(V_i) = 0$ *and given covariances.*

We are now in a position to understand the statement of Theorem 3.1. We begin by considering statement (i).

Almost sure convergence means that the probability $\mathbb{P}$(limit holds) $= 1$ in other words

$$\mathbb{P}(\lim_{n\to\infty} \frac{X_{ni}}{n} = 2^{-1-i}) = 1$$

So Let X_n be the number of leaves of a random recursive tree. Almost sure convergence means that for almost all sequences $(X_{i,n})$ there exists $N_{i,j}$ such that $|\frac{X_{i,n}}{n} - 0.5| < N_{i,j}$ for all $n \geq j$.

We must interpret this theorem as , given the space of random recursive trees 1 sample from this space is an entire random recursive tree. This is an example of the strong law of convergence i.e. a trajectory of values of X_{ni}/n will stay close to 2^{i-1} for any given trajectory.

This means that samples are points in the Weak Law for Large Numbers and entire trajectories in Strong Law for Large Numbers.